

VARSHA ASHOK KATARE

Email:katarevarsha256@gmail.com

Mobile: +91-7020074036

Career Objective:

To contribute my education and skills in a position with a growing and dynamic company which will utilize my skills and provide a supportive, challenging work environment.

Professional Summary:

- 3 + years of Extensive Experience in different domains in IT industry including Hadoop Ecosystem, Spark-Python API's, Cloud Services (AWS), NoSQL Database such as Hbase
- Hands on experience with Big Data core components and Eco System including Data Ingestion and Data Processing (PySpark, Hive, SQOOP, HDFS and MapReduce.)
- In depth understanding of Spark Architecture including SparkCore, Spark-SQL, DataFrames.
- Have a knowledge on Amazon web services mainly S3, EC2, Aws EMR, IAM Role, RDS.
- Uploaded and Processed terabytes of data from various structured and unstructured sources into HDFS using Sqoop, Spark.
- Capable of processing large sets of structured, semi-structured data.
- Very Good understanding of Partitioning, bucketing concepts in Hive and designed both Internal and External tables in Hive to optimize performance.
- Solved performance issues in Hive scripts with understanding of Joins, Group and Aggregation.
- Involving in Importing/Exporting data to/from RDBMS using Sqoop.
- Configuring various XMLs file such as Core-site.xml, Hdfs-site.xml, Mapredsite.xml, Yarn-site.xml.
- Experience in basic Linux Scripting. .
- Strong problem solving & technical skills coupled with clear decision-making.
- Hands-on Experience in mainframes and Hadoop eco-system.
- Good exposures in Requirement Analysis, Customization and user acceptance.

Education:

Done Masters's Degree in Computer science from University of Pune (Grade A)

Technology Proficiency:

- **Platforms:** Windows, Linux
- **Programming language:** Python.
- **Scripting Language:** Shell Script.
- **Big Data Ecosystem:** HDFS, Sqoop, Hive, SparkCore, Spark-SQL
- **IDE:** PyCharm
- **Cloud Technologies:** AWS (S3, EC2, RDS, IAM, & EMR Cluster), Glue
- **Relational database:** MySQL, Oracle, PostgreSQL.
- **Ticketing tool:** Jira.

Organizational Scan:

- **Organization:** Advent informatics Pvt. Ltd
- **Designation:** Data Engineer
- **Location:** Pune
- **Duration:** From 1-6-2019 to Till Date.

Professional Experience:

Project #1: Product Analysis and Marketing[PAAM] **Time Duration:** May 2019 – Mar 2020

Description:

Product Analysis and Marketing[PAAM] is Data Hub where analysis for data is done on AWS glue platform.

Roles & Responsibilities

- We used AWS Glue as cloud based ETL and Spark Glue job to process data.
- We used AWS S3 for Upstream and Downstream Data storage.
- Responsible for doing different transformation on big data like data cleaning, deduplication, Join etc.
- Set up different Glue jobs to dumps data into RAW, Cleansing and Process bucket. Thereafter created 5 different sub folders inside each 3 main buckets.
- Set up different scenario to write data using Pyspark Glue job if job run successful, failed, fetching errors, back up and processed.
- Monitoring the job using CloudWatch if any problem occurs it will helpful for trigger events.
- At the end uploading the output data to S3 Process Bucket which is final processed data ready for visualization.
- Analytical thinking about handling long running jobs to minimize time required for completion using optimization techniques.
- Hands-on experience with Amazon services like Glue, EC2, Amazon S3, Amazon RDS, EMR etc.

Project #2: Financial Intelligence Unit **Time Duration:** July 2020-Till Date

Description:

Financial Intelligence Unit deal with Banking domain.

Roles & Responsibilities:

- Understand the upstream source nature and work with data storage to handle historical, reference, and delta data using PySpark.

- Data migration from various sources such as S3 to Hbase, Hbase to Hive, Hive to S3 using Pyspark.
- Shared responsibility for administration of Hadoop, Hive.
- Copy All dependencies on EMR.
- Create Hbase table using phoenix Script.
- Written Spark job to Process data.
- Import Ref data from RDBMS to Hbase.
- Create Hive tables deployment to transformed data using hive DDL table.
- Working experience in handling Partitions, bucketing to optimize performance concepts in Hive and designing both Managed and External tables in Hive
- Responsible for doing different transformation on reference and delta data like data cleaning, deduplication, Join, etc. using Pyspark.
- To Process delta, use script Copy data into Hbase at data location.
- Customer can extract report from S3 and can use for their further use like visualization etc. And extract data
- To analyze and solve problems and put hot fixes on production.

Personal Details:

Full Name	:	Varsha Ashok Katare
Date of Birth	:	25/06/1990
Gender	:	Female
Marital Status	:	Married
Languages Known	:	English, Marathi & Hindi.
Nationality	:	Indian
Permanent address	:	Nandanvan housing society, RM-87/4, Bajaj Nagar, Aurangabad-431136.

I solemnly declare that all the information provided by me in this application is correct and complete to the best of my knowledge.

Date:

(Varsha Katare)